

Automatic Feature Learning Method for Detection of Retinal Landmarks

Baidaa Al-Bander, Waleed Al-Nuaimy, Majid A. Al-Tae, Ali Al-Ataby

Department of Electrical Engineering and Electronics
University of Liverpool, Liverpool, UK
e-mail: {hsbalban, wax, altaeem, aliataby}@liv.ac.uk

Yalin Zheng

Department of Eye and Vision Science
University of Liverpool, Liverpool, UK
e-mail: yalin.zheng@liv.ac.uk

Abstract— This paper presents an automatic deep learning method for the location detection of important retinal landmarks, the fovea and optic disc (OD) in digital fundus retinal images. The proposed method, which is based on deep convolutional neural networks (CNN) does not depend the visual appearance or anatomical features of the retinal landmarks. It comprises convolution, max-pooling, fully connected and dropout layers as well as an output layer. The CNN is trained using an existing dataset images along with their annotated locations of the foveal and OD centres. Performance of the network is evaluated using Root Mean Square Error (RMSE). The developed feature learning approach presents a promising system for retinal landmark detection.

Keywords— *Automatic feature learning; convolutional neural network; deep learning; retinal landmarks; automated grading*

I. INTRODUCTION

The human retina comprises several components, including the optic disc (OD), blood vessels, the fovea, and the macula. The OD appears as a bright yellowish region within color fundus images through which the blood vessels enter the eye. The macula is the center of the retina that is responsible for our central vision. The fovea is a small depression in the center of the macula. The location of the fovea center is about 2.5 ODDs from the optic disc center. The foveal radius is between 1/3 and 1/4 of the macula radius which is roughly equal to 1 ODD [1], [2].

Automatic detection of retinal anatomical structure from digital fundus images has been receiving increasing emphasis in the medical image processing community [3]–[7]. In particular, detecting location of the retinal landmarks played an important role in identifying retinal pathologies including glaucoma, diabetic maculopathy (DR), and age related macular degeneration (AMD). Importance of fovea detection is based on the fact that the nearer the center of the fovea a lesion is, the severer this lesion is. Moreover, center of OD can be used as a reference point for annotating other retinal structure. The OD can also be used as a starting point for blood vessels tracking [8].

Numerous methods have been reported in the literature [9]–[15] addressing various issues relevant to the visual appearance or anatomical features of OD and fovea such as Gegundez-Arias et al. [15] based on priori known anatomical features to detect the location of fovea. These methods that aimed at improving detection of the anatomical features of OD and fovea and, in particular, their positions involved. Furthermore, various machine-learning analytics were proposed, including; k-nearest neighbor [16],

ensemble based methods [17], [18]. In [17], the authors organized the most recent OD detectors into an ensemble and complex framework to maximize the accuracy of OD detection. However, performance of these analytics was highly dependent on the quality of the features extraction method.

For a long time, the features have been extracted manually through adopting various feature-engineering approaches. However, the existing generic guidelines for extracting features from different datasets cannot fulfill the requirements of different datasets. For example, the features that are usable for one dataset are often not usable for other datasets. Therefore, the search for new algorithms that are capable of learning features automatically has become a key requirement for developing more accurate machine learning analytics [19].

Nowadays, a new area in machine learning analytics that is called deep neural networks (DNNs) has been emerged. Unlike the conventional ANN, the layers of DNN are not fully connected and can learn to recognize highly complex non-linear features in its input. In addition, DNNs are also based on graphical processing units (GPUs) that have become the platform of choice for training large and complex DNN-based machine learning in both local and distributed application levels. Various deep learning architectures such as convolutional neural network (CNN), deep belief network (DBN), and restricted Boltzmann machine (RBM) have reported and applied to various applications including computer vision [20], automatic speech recognition [21], natural language processing [22], and bioinformatics [23] where they have been shown to produce state-of-the-art results on various tasks.

In this paper, a new automatic feature learning method is proposed to detect the OD and fovea locations simultaneously (joint OD-fovea detection). The proposed method that is based on deep CNN does not depend the visual appearance or anatomical features of the landmarks under study. Furthermore, unlike traditional machine learning and feature engineering algorithms, the hierarchical extracted features of the proposed method are automatically learned from the dataset and thus address the challenges of different datasets on individual bases.

The remainder of this paper is organized as follows. Section II describes the convolutional neural network. The dataset adopted in this study is described in Section III along with the proposed feature learning and detection method. Next, performance evaluation of the developed algorithm for retinal landmarks detection is presented and discussed in Section IV. Finally, the presented work is concluded in Section V.

II. CONVOLUTIONAL NEURAL NETWORK

In contrast to conventional shallow classifiers, such as neural networks and support vector machines, for which a feature extraction step is essential, hierarchies of significant features are learnt by deep learning algorithms directly from the raw input data.

A block diagram representation for the proposed convolutional neural network architecture is shown in Fig. 1. It comprises three convolutional layers, three pooling layers and four dropout layers followed by fully connected layers and an output classification/regression layer, as illustrated. These layers are described briefly as follows.

1) *Convolutional layer*: The convolutional layer comprises a sequence of filters to perform a 2D convolution on the input images. The convolutional filter has dimensions much smaller than the image it works on connecting to local regions in the input images with weights shared between all filters. The output of each input map in the images (feature map) is determined by summing the responses over the whole input map [24].

2) *Max-pooling layer*: The feature map resulting from a convolution layer is usually down-sampled by non-overlapped square regions, where the size of this square region is a hyper-parameter that could be empirically adjusted by the user. This square region (window) is moved over the feature map: each time the highest activation in this local region is picked out while other values are omitted. The motivation and intuition behind this layer is to speed up convergence by decreasing the number of parameters to be learnt and the computation time in the neural network [25].

3) *Dropout layer*: A dropout layer [26] is an effective regularization technique that helps to reduce the overfitting during training. The overfitting problem can be reduced by randomly deleting the node in the layer with a certain probability at each training step. A deleted unit will not participate in forward propagation and backpropagation in the training stage that is achieved using the stochastic gradient descent (SGD) algorithm. All of deleted units are re-enabled at testing stage by multiplying them with one minus the probability of dropping out.

4) *Fully Connected layer*: It usually represents the top layers of deep neural network architecture. Each neuron in this layer is connected completely to all of the neurons in the previous layer and the weights of these connection are specific to each neuron. The number of neurons in the output layer which is fully connected layer represents a one neuron per class (one neuron for each coordinate of F and OD in this work)

5) *Activation functions*: Deep networks usually comprise convolution filters followed by a non-linear activation function after each layer. Rectified activation function is used after all convolution layers in our proposed network. A unit employing the rectifier is called a Rectified Linear Unit (RELU) [27]. The rectified function is defined by

$$\varphi: x \rightarrow \max(0, x). \quad (1)$$

Architecture of the implemented convolutional network is described in Table I. Last column shows the size of filters, window size of max-pooling, and probability of dropping a node in each layer.

TABLE I. DEEP NURAL NETWORK ARCHITECTURE

Name	Size	No. of outputs	Filters#	Size of Filter, maxpool.,Probability
input	1×256×256	65536	-	-
conv1	16×254×254	1032256	16	filter size = (3,3)
pool1	16×127×127	258064	-	maxpool size = (2,2)
dropout1	-	-	-	dropout1_p = 0.05
conv2	32×126×126	508032	32	filter size = (2,2)
pool2	32×63×63	127008	-	maxpool size = (2,2)
dropout2	-	-	-	dropout2_p = 0.1
conv3	32×62×62	246016	64	filter size = (2,2)
pool3	32×31×31	61504	-	maxpool size = (2,2)
dropout3	-	-	-	dropout3_p = 0.2
FC	250	250	-	-
dropout4	-	-	-	dropout3_p = 0.3
FC	250	250	-	-
output	4	4	-	-

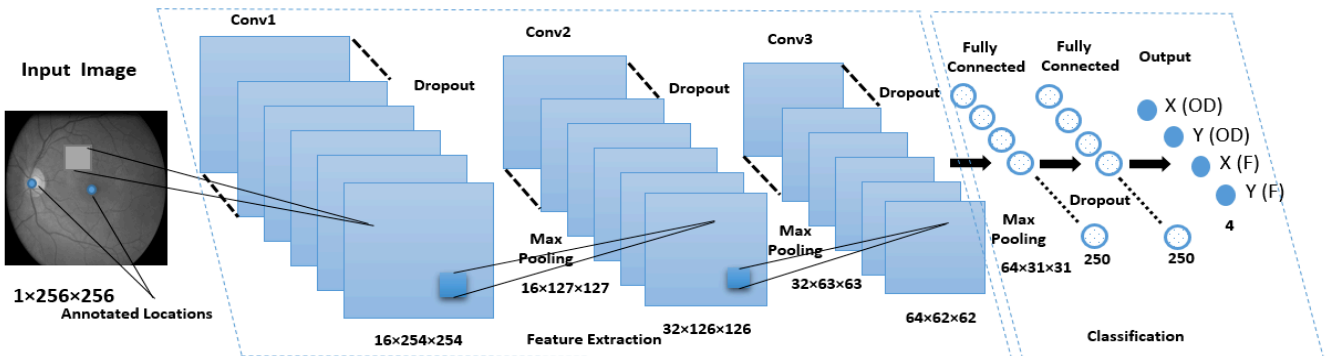


Figure 1. Convolutional neural network architecture

III. MATERIALS AND METHOD

A. Dataset

The ARIA dataset [28] is adopted in this study. It comprises 120 images evenly divided between normal and diabetic images. These images were captured using Zeiss FF450+ fundus camera with 50-degree field of view and a resolution of 576×768 pixels. A trained image-analysis expert had marked up the border of the optic disc as well as the fovea location in the original dataset. However, the coordinates of the optic disc and fovea centers are unknown.

B. Method

The proposed feature learning and detection method consists of three main phases: pre-processing, CNN training, and a testing and evaluation stage as shown in the block diagram of Fig. 2. These stages are outlines as follows.

1) *Pre-processing*: At this stage, the images are prepared for training. The optic disc and foveal center point coordinates are located and calculated depending on ground truth images, as shown in Fig. 3. This is achieved by finding the two objects (border of OD and the fovea object) in the provided ground truth images. After that, the centroid of these objects are identified. For the purpose of this study, the images were resized to 256×256 pixels while the annotated center point coordinates of both the OD and fovea were scaled accordingly. These annotated locations together with the images were used to train and evaluate the performance of the implemented networks.

2) *Convolutional neural network*: As described in Section II, architecture of the CNN comprises convolutional filters which are responsible of extracting the features, subsampling layers to reduce the number of parameters, dropout layers to decrease the overfitting, and the output layer that represent the axis coordinates of OD and F. Since detecting the landmarks' location is a regression task, mean

square error (MSE) is the objective function that should be minimized in the output layer.

3) *Network training*: The dataset was randomly divided into 90 images for training and validation (20% of training data is a validation data), and the remaining 30 for testing. Furthermore, detecting the center points of the OD and F is a regression task, therefore; the gray scale images were fed to the convolutional network. It is not necessary to use the colors of the images in this regression task because it just adds extra complexity. Moreover, the pixel values of images were scaled between $[0, 1]$ rather than $[0, 255]$ and the annotated location of OD and F values were scaled between $[-1, 1]$. After training stage, the trained network is used to test the unseen images to evaluate the system generalization performance.

4) *Test and evaluation*: Once the network is trained, the convolution neural network will be able to perform the prediction on the test data set. In this stage, the performance of our implemented network is assessed.

IV. RESULTS AND DISCUSSION

All experiments are conducted on a Linux Mint machine with 16GB RAM, Intel Xeon 3.50GHz processor and NVIDIA Quadro K2200 GPU with 640 CUDA parallel-processing cores. We use Lasagne [29] Python deep learning library built on the top of Theano [30] to implement and train the convolutional neural networks. To train the networks by updating the weights, SGD with a momentum optimization algorithm having learning rate (0.01) and momentum parameter (0.9) is used. Weight initialization method proposed in [31] is considered to initialize the weights of kernels in the implemented network and the network is trained with 500 epoch.

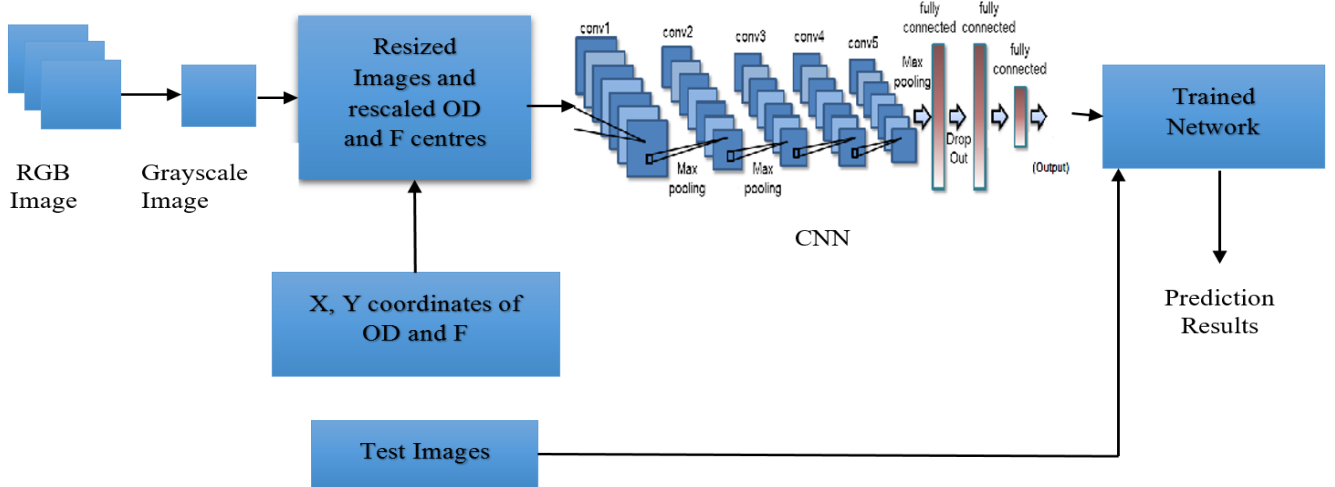


Figure 2. Stages of the proposed feature learning and detection method

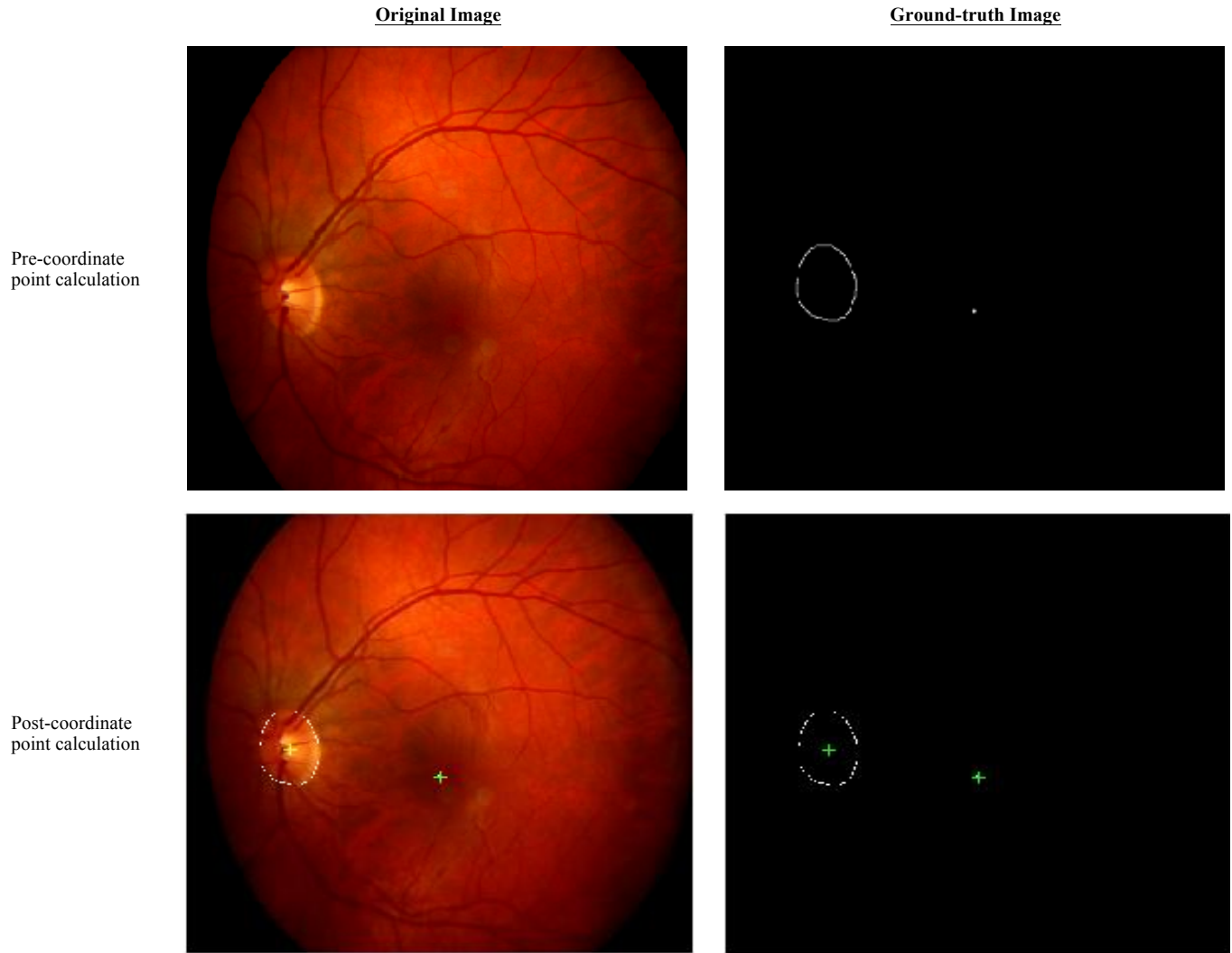


Figure 3. Preparation of dataset Images

Learning performance of the implemented network has been monitored during the training stage by plotting the learning curves for both training and validation sets by determining the loss in terms of RMSE as shown in Fig. 4. As illustrated, the model suffers from slight overfitting (high variance) where the gap between the training and validation error indicates the amount of overfitting. Also, we can notice from the figure that the training error is much less than validation error. More examples are therefore needed to improve the validation error. The reason behind this overfitting problem is the CCNs typically need larger training dataset to extract a more efficient set of features. In addition to adding more data, data augmentation technique allows for artificially increasing the number of training examples through applying transformation, adding noise to the images and others.

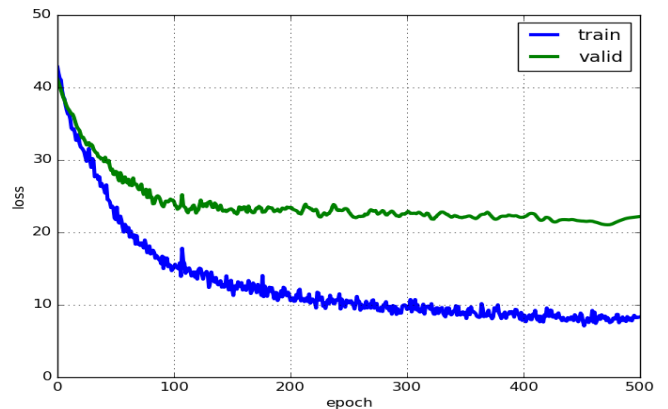


Figure 4. Learning performance of the conventional neural network

The implemented network has been tested on the test set as shown in Fig. 5. As illustrated in the images of this figure, the proposed CNN has successfully extracted the features of interest (i.e. the optic disc and fovea). For the purpose of performance comparison, the methods reported in the literature had used $1R$ criteria to evaluate and assess their retinal landmark detection methodologies where R refers to the OD radius. The distance between the ground truth and the obtained location of the OD or foveal center is compared with the $1R$ -value in each image to determine the validity of the location determined by the automated detection methods. Many studies [4]-[6],[32] have established that the obtained detection of the landmark center is correct if the Euclidean distances to the ground truth centers is within half the OD diameter. Moreover, there is no fix rule or formula to calculate $1R$ -value where the aforementioned studies used different OD radii in their studies even though they worked on the same dataset. As a result of the mentioned reasons, a strategy should be proposed and considered to determine the OD radius in the

data set that is used in our work in order to be able to compare our obtained results with other methodologies in the literature. Furthermore, for training, adaptive values for both learning rate and momentum parameter can be used rather than the constant values to speed up the training process. In addition to that, the poor quality images could be further enhanced using image enhancement techniques as a preprocessing step.

V. CONCLUSIONS

In this paper, an automated method to locate the most important anatomical landmarks in fundus images has been presented. The benefits of features learned automatically from the retinal images through the stacked layers of convolution filters have been designed, realized and evaluated. The developed method, which is based on deep convolutional neural network, has proved to be efficient in feature learning and identification of retinal landmarks of interest.

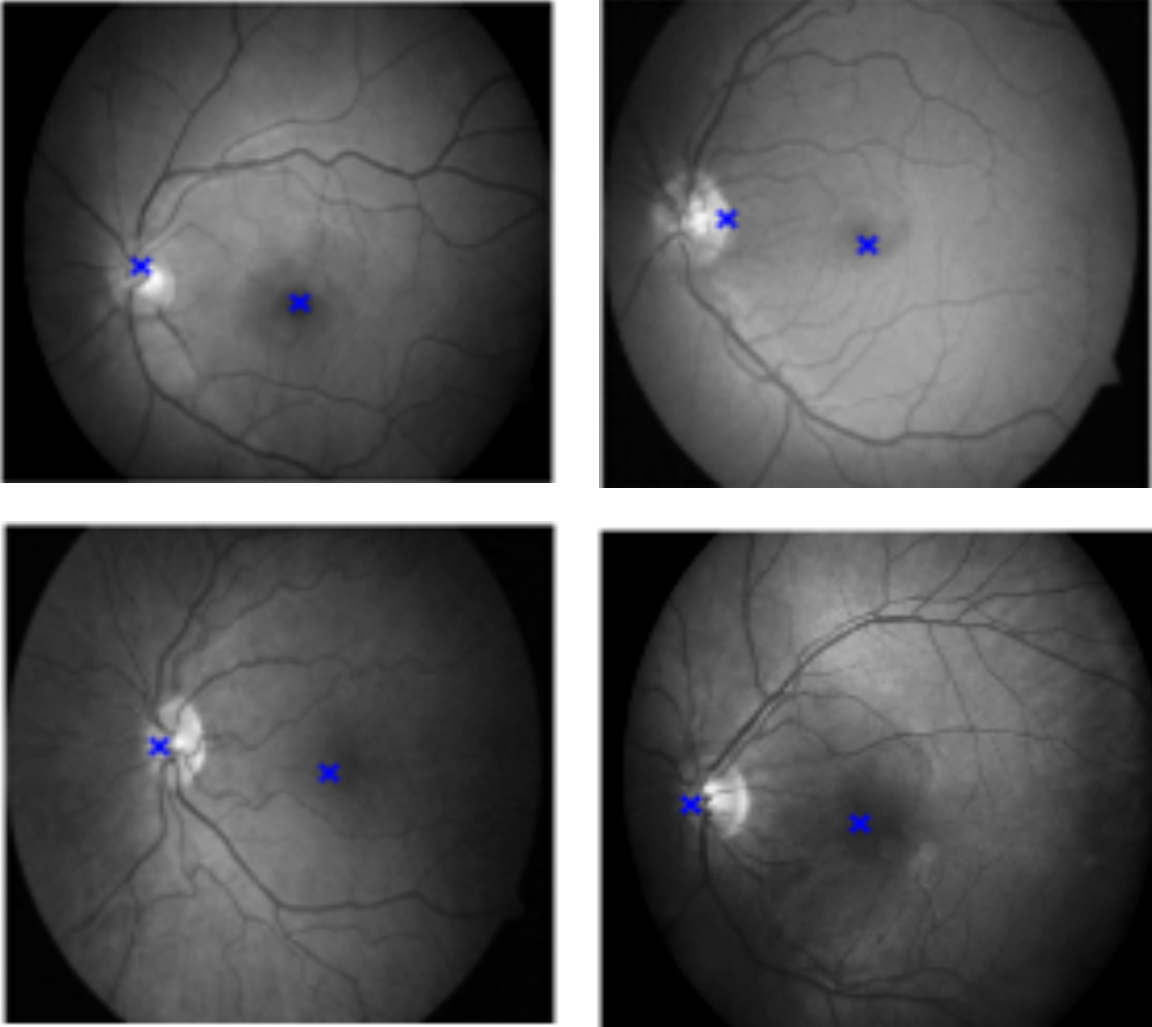


Figure 5. Examples of joint OD-Fovea prediction results on. The left cross in each image represents the OD centers and the right represents the fovea center

The obtained results demonstrated that the proposed feature learning method is promising in identifying locations of anatomical structure efficiently and accurately. Implementing multiple layers of learning filters has been proved to be superior to current approaches in extracting the features automatically from the data and thus solved the issues related to the workload in extracting the features manually from various types of datasets. The proposed work, however, is still open for further improvements including (i) increasing the size of dataset by applying a more efficient data augmentation technique, (ii) utilizing an adaptive value for both the learning rate and momentum, (iii) using more advanced $1R$ criteria that depends dynamic optic disc radius rather than using a fixed radius, and (iv) using multiple performance evaluation metrics to assess the detection accuracy of the landmarks. These improvements are currently part of the authors' ongoing research.

REFERENCES

- [1] M. Niemeijer, M. D. Abràmoff, and B. Van Ginneken, "Segmentation of the optic disc, macula and vascular arch in fundus photographs", *Medical Imaging, IEEE Transactions on*, 26(1), 2007, 116-127.
- [2] X. Zhu, R. M. Rangayyan, and A. L. Ells, "Detection of the optic nerve head in fundus images of the retina using the Hough transform for circles", *Journal of Digital Imaging*, 23(3), 2010, 332-341.
- [3] R. J. Qureshi, L. Kovacs, B. Harangi, B. Nagy, T. Peto, and A. Hajdu, "Combining algorithms for automatic detection of optic disc and macula in fundus images", *Computer Vision and Image Understanding*, 116(1), 2012, 138-145.
- [4] A. Giachetti, L. Ballerini, E. Trucco, and P. J. Wilson, "The use of radial symmetry to localize retinal landmarks", *Computerized Medical Imaging and Graphics*, 37(5), 2013, 369-376.
- [5] M. E. Gegundez-Arias, D. Marin, J. M. Bravo, and A. Suero, "Locating the fovea center position in digital fundus images using thresholding and feature extraction techniques", *Computerized Medical Imaging and Graphics*, 37(5), 2013, 386-393.
- [6] A. Aquino, "Establishing the macular grading grid by means of fovea center detection using anatomical-based and visual-based features", *Computers in Biology and Medicine*, 55, 2014, 61-73.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", In *Advances in Neural Information Processing Systems*, 2012, (pp. 1097-1105).
- [8] L. Gagnon, M. Lalonde, M. Beaulieu, and M. C. Boucher, "Procedure to detect anatomical structures in optical fundus images", In *Medical Imaging 2001* (pp. 1218-1225). International Society for Optics and Photonics.
- [9] S. Lu, "Accurate and efficient optic disc detection and segmentation by a circular transformation", *Medical Imaging, IEEE Transactions on*, 30(12), 2011, 2126-2133.
- [10] A. Hoover, and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels", *Medical Imaging, IEEE Transactions on*, 2003, 22(8), 951-958.
- [11] M. Foracchia, E. Grisan, and A. Ruggeri, "Detection of optic disc in retinal images by means of a geometrical model of vessel structure", *Medical Imaging, IEEE Transactions on*, 23(10), 2004, 1189-1195.
- [12] A. D. Fleming, K. A. Goatman, S. Philip, J. A. Olson, and P. F. Sharp, "Automatic detection of retinal anatomy to assist diabetic retinopathy screening", *Physics in Medicine and Biology*, 52(2), 2006, 331.
- [13] K. W. Tobin, E. Chaum, V. P. Govindasamy, & T. P. Karnowski, "Detection of anatomic structures in human retinal imagery", *Medical Imaging, IEEE Transactions on*, 26(12), 2007 1729-1739.
- [14] A. Yousif, A. Z. Ghalwash, and A. Ghoneim, "Optic disc detection from normalized digital fundus images by means of a vessels' direction matched filter", *Medical Imaging, IEEE Transactions on*, 27(1), 2008, 11-18.
- [15] M. E. Gegundez-Arias, D. Marin, J. M., Bravo, and A. Suero, "Locating the fovea center position in digital fundus images using thresholding and feature extraction techniques", *Computerized Medical Imaging and Graphics*, 37(5), 2013, 386-393.
- [16] M. Niemeijer, M. D. Abràmoff, and B. Van Ginneken, "Fast detection of the optic disc and fovea in color fundus photographs", *Medical Image Analysis*, 13(6), 2009, 859-870.
- [17] B. Harangi, and A. Hajdu, "Detection of the optic disc in fundus images by combining probability models", *Computers in Biology and Medicine*, 65, 2015, 10-24.
- [18] R. J. Qureshi, L. Kovacs, B. Harangi, B. Nagy, T. Peto, and A. Hajdu, "Combining algorithms for automatic detection of optic disc and macula in fundus images", *Computer Vision and Image Understanding*, 116(1), 2012, 138-145.
- [19] I. Guyon, S. Gunn, M. Nikravesh, & L. A. Zadeh, "Feature extraction: foundations and applications", (Vol. 207), 2008, Springer.
- [20] S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. Kruthiventi, and V. B. Radhakrishnan, "A taxonomy of Deep Convolutional Neural Nets for Computer Vision", *Frontiers in Robotics and AI*, 2, 36, 2015.
- [21] O. Abdel-Hamid, A. R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition", *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 22(10), 2014, 1533-1545.
- [22] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent Convolutional Neural Networks for Text Classification", In *AAAI*, 2015, (pp. 2267-2273).
- [23] T. Zeng, R. Li, R. Mukkamala, J. Ye, and S. Ji, "Deep convolutional neural networks for annotating gene expression patterns in the mouse brain", *BMC bioinformatics*, 16(1), 2015.
- [24] Y. LeCun, L. Bottou, Y. Bengio, & P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, 86(11), 1998, 2278-2324.
- [25] M. A. Ranzato, F. J. Huang, Y. L. Boureau, and Y. LeCun, "Unsupervised learning of invariant feature hierarchies with applications to object recognition", In *Computer Vision and Pattern Recognition*, 2007. *CVPR'07. IEEE Conference on* (pp. 1-8).
- [26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting", *The Journal of Machine Learning Research*, 15(1), 2014, 1929-1958.
- [27] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks", In *International Conference on Artificial Intelligence and Statistics*, 2011, pp. 315-323.
- [28] Y. Zheng, M. H. A. Hijazi, and F. Coenen, "Automated "disease/no disease" grading of age-related macular degeneration by an image mining approach", *Investigative Ophthalmology & Visual Science*, 53(13), 2012, 8310-8318.
- [29] Lasagne, available online: <https://github.com/Lasagne/Lasagne> (Accessed on 28 June 2016).
- [30] Theano, available online: <https://github.com/Theano/Theano>, (Accessed on 28 May 2016).
- [31] X. Glorot, and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks", In *International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249-256.
- [32] H. Yu, E. S. Barriga, C. Agurto, S. Echegaray, M. S. Pattichis, W. Bauman, and P. Soliz, "Fast localization and segmentation of optic disk in retinal images using directional matched filtering and level sets", *Information Technology in Biomedicine, IEEE Transactions on*, 16(4), 2012, 644-657.